A note on upper bounds for forecast-value-added relative to naive forecasts

Paul Goodwin

p.goodwin@bath.ac.uk

Fotios Petropoulos

f.petropoulos@bath.ac.uk

The Management School

University of Bath

BA2 7AY

United Kingdom

Rob J. Hyndman

Rob.Hyndman@monash.edu

Department of Econometrics and Business Statistics

Monash University

Wellington Rd,

Clayton VIC 3800,

Australia

ABSTRACT

In Forecast Value Added (FVA) analysis the accuracy of relatively sophisticated forecasting methods is compared to that of naïve 1 forecasts to see if the extra costs and effort of implementing them is justified. In this note we derive a ratio that indicates the upper bound of a forecasting method's accuracy relative to naïve 1 forecasts when the mean squared error is used to measure one-period-ahead accuracy. The ratio is applicable when a series is stationary or when its first differences are stationary. Formulae for the ratio are presented for several exemplar time series processes.

1.0 Introduction

In recent years Forecast Value Added (FVA) analysis has being attracting attention in the forecasting practitioner literature (e.g. Gilliland, 2011, Morlidge, 2014). The idea is simple. To assess whether the costs and effort of applying sophisticated forecasting methods are justified their accuracy is compared to that of a relatively simple method. Usually 'naïve 1' forecasts act as the simple benchmark method. A naïve 1, or no-change, forecast is equal to the latest observation in a time series (hereafter they will be referred to as "naïve" forecasts) (McLaughlin, 1983). To assist this process of comparison a number of error measures have been designed to compare directly the accuracy of a given method with that of naïve forecasts. These include Theil's U^2 (Theil, 1971), the relative absolute error (RAE) (Armstrong and Collopy, 1992) and the mean absolute scaled error (MASE) (Hyndman and Koehler, 2006). When a time series is a random walk, the naive forecasts will be optimal and, because a number of processes in areas like financial and stock market forecasting can be approximated by random walks, naïve forecasts are sometimes difficult to beat. The same may be true in many product demand forecasting contexts. For example, in an analysis of over 300,000 forecasts in eight companies, Morlidge (2014) found that over half were less accurate than naïve forecasts. Of course, results like this may, at least in part, reflect the use of inappropriate forecasting methods or unwarranted judgmental adjustments to forecasts rather than the inherent unpredictability of the product time series.

This raises the question of how much improvement in accuracy one can expect over naïve forecasts in different situations and, in particular, where the upper limit of potential improvements lies. In this paper a ratio is obtained to identify this limit under conditions where the values in a time series follow a stochastic process and the noise is homoscedastic.

The use of the ratio is limited to series which are stationary or are stationary when first differences are taken. For other series, it would not usually be sensible to use naïve forecasts as a comparator. For example, naïve forecasts could not be expected to perform well in a series where the underlying signal takes on values of 1, 2, 4, 9, 16, 25 and so on (i.e., a series that is stationary only when second differences are taken). The ratio also assumes that forecasts are being made for one period ahead and that it is appropriate to measure accuracy using the mean squared error (MSE) where:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(y_t - f_t)^2}{n}$$

and y_t = actual observation at time t, f_t = the forecast for time t and n = the number of forecasts. The MSE assumes that a quadratic loss function is applicable.

2.0 Derivation of the ratio for ARIMA($p, 0, q$) models

First, consider the case of an ARIMA($p, 0, q$) model, where p is the order of the autoregressive part of the model and q is the order of the moving average part. The model is:

$$\phi(B)y_t = c + \theta(B)e_t,$$

where $e_t \sim$ iid $N(0, \sigma^2)$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ and B is the backshift operator.

Hence:

$$y_t = \frac{c}{\phi(1)} + \psi(B)e_t,$$

where $\psi(B) = \theta(B)/\phi(B) = \sum_{i=0}^{\infty} \psi_i B^i$ and $\psi_0 = 1$.

Then the first difference is given by

$$(1 - B)y_t = (1 - B)\psi(B)e_t,$$

and the MSE of a naive forecast is

$$V = \mathrm{E}[(1 - B)y_t]^2 = \mathrm{E}[(1 - B)\psi(B)e_t]^2 = \sigma^2 \left[1 + \sum_{i=1}^{\infty}(\psi_i - \psi_{i-1})^2\right].$$

The MSE of the optimal forecast is σ^2. So the Ratio is given by

$$\text{Ratio} = \frac{\sigma^2}{V} = \frac{1}{1+\sum_{i=1}^{\infty}(\psi_i - \psi_{i-1})^2}. \tag{1}$$

2.0 Derivation of the ratio for ARIMA($p, 1, q$) models

Now, consider the case of an ARIMA($p, 1, q$) model:

$\phi(B)(1-B)y_t = c + \theta(B)e_t$.

Then

$$(1-B)y_t = \frac{c}{\phi(1)} + \psi(B)e_t,$$

and the MSE of a naive forecast is

$$V = \mathrm{E}[(1-B)y_t]^2 = \frac{c^2}{\phi^2(1)} + \mathrm{E}[\psi(B)e_t]^2 = \frac{c^2}{\phi^2(1)} + \sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

The MSE of the optimal forecast is σ^2. So the Ratio is given by

$$\text{Ratio} = \frac{\sigma^2}{V} = \frac{1}{\frac{c^2}{\sigma^2\phi^2(1)} + \sum_{i=0}^{\infty}\psi_i^2}. \tag{2}$$

3.0 Values of the ratio for exemplar series

The values of the ratio for a number of exemplar time series processes are shown in Table 1 In practice these ratios can provide guidance on what might be expected if a forecasting model is being applied appropriately under ideal conditions on a large sample of time series observations. Note that if the series consists only of simply random variation around a fixed mean, then $\psi_0 = 1$ and $\psi_i = 0$ for i ≥ 1 so the ratio in (1) is 0.5. Hence choosing to use naïve forecasts in these circumstances is likely to be costly resulting in a potential doubling of the MSE. From the ratios in Table 1 it can be seen that naïve forecasts will perform relatively well where a series exhibits high positive first order autocorrelation since successive observations will tend to be on the same side of the mean level and hence relatively close. The opposite is true for negative autocorrelation. For example, when ϕ_1 is 0.7 in the

ARIMA(1,0,0) model, the first-order autocorrelation is also 0.7, and the maximum reduction that can be achieved over the naïve forecasts' MSE is only 15%. When ϕ_1 = -0.7 it is 85%.

For the ARIMA(0,1,1) series, when the intercept is zero, simple exponential smoothing with a smoothing constant of 1-θ_1 is optimal. In practice values of smoothing constants between 0.1 and 0.3 are often employed (Gardner, 1985). Assuming that when these values are used they are optimal, and that simple exponential smoothing is appropriate, the ratio shows that MSE reductions over naïve forecasts of between about 45% and 33%, can be achieved. However, in general for an ARIMA(0,1,1) series, simple exponential smoothing can only reduce the MSE of naïve forecasts by a maximum of 50% -the limit applies when the optimal smoothing constant is zero indicating a series with a constant mean.

ARIMA (0,0,1) series

Model: $y_t = c - \theta_1 e_{t-1} + e_t$ where the condition for invertibility is -1< θ_1 <1

So $\psi_1 = -\theta_1$, $\psi_i = 0$ for $i \geq 2$ in (1), and

Ratio = $\dfrac{1}{2(1+\theta_1+\theta_1^2)}$

ARIMA (0,0,2) series

Model: $y_t = c - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$

where the conditions for invertibility are: -1< θ_2 <1; $\theta_2 + \theta_1$ <1; $\theta_2 - \theta_1$ <1

So $\psi_1 = -\theta_1$, $\psi_2 = -\theta_2$, $\psi_i = 0$ for $i \geq 3$ in (1) and

Ratio = $\dfrac{1}{2(1+\theta_1+\theta_1^2+\theta_2^2-\theta_1\theta_2)}$

ARIMA (1,0,0) series

Model: $y_t = c + \phi_1 y_{t-1} + e_t$

where the condition for stationarity is -1< ϕ_1 <1

So $\psi_i = \phi_1^i$, $\psi_i - \psi_{i-1} = \phi_1^{i-1}(\phi_1 - 1)$ for $i \geq 1$ in (1), and

Ratio: $\dfrac{1+\phi_1}{2}$

ARIMA (0,1,1) series

Model: (1-B) $y_t = c - \theta_1 e_{t-1} + e_t$

So $\phi(1) = 1$, $\psi_1 = -\theta_1$, and $\psi_i = 0$, $i \geq 2$ in (2), and

Ratio = $\dfrac{\sigma^2}{\sigma^2(1+\theta_1^2)+c^2}$ which simplifies to $\dfrac{1}{1+\theta_1^2}$ if $c = 0$

ARIMA (1,1,0) series

Model: $(1-\phi_1 B)(1-B)y_t = c + e_t$

So $\phi(1) = 1 - \phi_1$, $\psi_i = \phi_1^i$ in (2), and

Ratio: $\dfrac{\sigma^2}{\sigma^2\left[\dfrac{1}{1-\phi_1^2}\right]+\left(\dfrac{c}{1-\phi_1}\right)^2}$

which simplifies to $1-\phi_1^2$ when c =0.

Table 1 Ratios for exemplar series

References

Armstrong JS and Collopy F (1992). Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting* 8: 69-80.

Gardner ES (1985). Exponential smoothing - the state of the art. *Journal of Forecasting* 4: 1-28.

Gilliland, M. (2011). Value added analysis: Business forecasting effectiveness, *Analytics*, July/August.

Hyndman RJ and Koehler AB (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22: 679-688.

McLaughlin RL (1983). Forecasting models - sophisticated or naive. *Journal of Forecasting* 2: 274-&.

Morlidge, S. (2014). Using relative error metrics to improve forecasting quality in the supply chain. *Foresight: The International Journal of Applied Forecasting* 34: 39-46.

Theil H (1971). *Applied Economic Forecasting*. North-Holland Publishing Company: Amsterdam.